Corrigé : Estimation Paramétrique

Exercice 1

1. Bernouilli $\mathcal{B}(p)$; $\theta = p \in [0, 1]$.
   $\mathbb{E}(X_1) = p$, donc $\hat{p}_n^{EMM} = \overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. C'est un estimateur sans biais (ESB) de p car $\mathbb{E}(\hat{p}_n^{EMM}) = p$. $\mathbb{P}(X_i = x_i) = p$, si $x_i = 1$, et $1 - p$ si $x_i = 0$. Donc
   $L = \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = p^{\sum x_i}(1-p)^{n-\sum x_i}$.
   $logL = (\sum x_i)\log p + (n - \sum x_i)\log(1-p)$.
   $\frac{\partial}{\partial p} logL = \frac{1}{p(1-p)}(\sum x_i - np) = 0 \Longrightarrow \hat{p}_n^{EMV} = \frac{1}{n}\sum_{i=1}^n X_i = \hat{p}_n^{EMM} := \hat{p}_n$.
   $\mathbb{V}(\hat{p}_n) = \frac{\mathbb{V}(X_1)}{n} = \frac{p(1-p)}{n}$.
   La quantité d'information de Fisher vaut : $I_n(p) = \mathbb{V}(\frac{\partial}{\partial p} logL) = \mathbb{V}(\frac{1}{p(1-p)}(\sum x_i - np)) = \frac{n}{p(1-p)}$.
   Ainsi, $\mathbb{V}(\hat{p}_n) = \frac{1}{I_n(p)}$=borne FDCR, donc $\hat{p}_n$ est optimal, car ESB et de variance minimale (efficace). Il n'est pas normal à n fixé, mais asymptotiquement normal (quand $n \to \infty$) par le TCL.
2. Exponentielle $\mathcal{E}(\lambda)$; $\theta = \lambda > 0$.
   On a $\mathbb{E}(X_1) = \frac{1}{\lambda}$, donc $\hat{\lambda}_n^{EMM} = \frac{1}{\overline{X}_n} = \frac{n}{\sum_{i=1}^n X_i}$.
   On peut montrer (en utilisant que $\sum_{i=1}^n X_i$ suit la loi Gamma $G(n, \lambda)$) que $\mathbb{E}(\hat{\lambda}_n^{EMM}) = \frac{n}{n-1}\lambda \neq \lambda$, donc c'est un estimateur biaisé (mais asymptotiquement sans biais).
   En écrivant $L = f_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n) = \prod_{i=1}^n \lambda \exp(-\lambda x_i)$, puis en faisant $\frac{\partial}{\partial \lambda} logL = 0$, on obtient $\hat{\lambda}_n^{EMV} = \hat{\lambda}_n^{EMM} = \frac{n}{\sum_{i=1}^n X_i}$.
3. Exponentielle $\mathcal{E}(\lambda)$; $\theta = \frac{1}{\lambda} > 0$.
   $\hat{\theta}_n^{EMM} = \hat{\theta}_n^{EMV} = \overline{X}_n$ devient simplement un ESB de θ.
4. Normale $\mathcal{N}(m, \sigma^2)$; $\theta = (m, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+$.
   $(m = \mathbb{E}(X); \sigma^2 = \mathbb{E}[(X - \mathbb{E}(X))^2])$, donc $(\hat{m}_n^{EMM} = \overline{X}_n; \hat{\sigma^2}_n^{EMM} = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2 := S_n^2)$ est un estimateur par méthode des moments de θ.
   En écrivant $L = f_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n) = \prod_{i=1}^n f_{X_1}(x_i)$ (avec $f_{X_1}(x_i) = \frac{1}{\sigma\sqrt{2\pi}}\exp(-\frac{(x_i-m)^2}{2\sigma^2})$), puis en faisant $\frac{\partial}{\partial m} logL = 0$ et $\frac{\partial}{\partial(sigma^2)} logL = 0$, on obtient les mêmes estimateurs par maximum de vraisemblance.
   $\hat{m}_n$ est un ESB de m, mais $\hat{\sigma^2}_n = S_n^2$ est biaisé : $\mathbb{E}(S_n^2) = \frac{n-1}{n}\sigma^2$. C'est pourquoi on utilise plutôt l'estimateur $S_n'^2 := \frac{n}{n-1}S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$ qui est alors un ESB de σ^2.
5. Gamma $\Gamma(\alpha, \lambda)$; $\theta = (\alpha, \lambda) \in \mathbb{R}_+^* \times \mathbb{R}_+^*$.
   On sait pour la loi Gamma que $\mathbb{E}(X) = \frac{\alpha}{\lambda}$ et $\mathbb{V}(X) = \frac{\alpha}{\lambda^2}$. Donc :
   $\alpha = \frac{(\mathbb{E}(X))^2}{\mathbb{V}(X)}$ et $\lambda = \frac{\mathbb{E}(X)}{\mathbb{V}(X)}$. Ainsi :
   $\hat{\alpha}_n^{EMM} = \frac{(\overline{X}_n)^2}{S_n^2}$ et $\hat{\lambda}_n^{EMM} = \frac{\overline{X}_n}{S_n^2}$. Ils sont biaisés.

En écrivant $L = f_{(X_1,\ldots,X_n)}(x_1,\ldots,x_n)$, puis en faisant $\frac{\partial}{\partial\alpha} logL = 0$ et $\frac{\partial}{\partial\lambda} logL = 0$, on obtient un système de deux équations, dont la résolution n'est pas explicite, et donc la solution est différente de $(\hat{\alpha}_n^{EMM}, \hat{\lambda}_n^{EMM})$. On a ainsi un contre-exemple où l'estimateur par max. de vraisemblance (EMV) est différent de celui par méthode des moments (EMM).

Exercice 2

Un estimateur linéaire et sans biais de θ s'écrit sous la forme $\hat{\theta}_n = \sum_{i=1}^n a_i X_i$ avec $\sum_{i=1}^n a_i = 1$. Comme les variables aléatoires sont indépendantes, on a $\mathbb{V}(\hat{\theta}_n) = \sum_{i=1}^n a_i^2\sigma^2$, avec $\sigma^2 = \mathbb{V}(X_1)$. D'après l'inégalité de Cauchy-Schwarz on a $(\sum_{i=1}^n a_i^2)(\sum_{i=1}^n \frac{1}{n^2}) \geq (\sum_{i=1}^n \frac{a_i}{n})^2$ Puisque $\sum_{i=1}^n a_i = 1$, on déduit que $\sum_{i=1}^n a_i^2 \geq \frac{1}{n}$ avec égalité si et seulement si $a_i = \frac{1}{n}$ pour tout $i \in \{1,\ \ldots,\ n\}$. L'estimateur de variance minimale dans la classe des estimateurs linéaires et sans biais est donc la moyenne empirique $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$.

Exercice 3

1. La loi de X_i est la loi $\mathcal{N}(m_i + \alpha,\ \sigma^2)$. La log-vraisemblance s'écrit :

$$\log(L_n(x_1,\ \ldots,\ x_n;\alpha)) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - m_i - \alpha)^2$$

On obtient :

$$\frac{\partial}{\partial\alpha}\log(L_n(x_1,\ \ldots,\ x_n;\alpha)) = \frac{1}{\sigma^2}\sum_{i=1}^n (x_i - m_i - \alpha)\ ,$$

et donc :

$$\frac{\partial}{\partial\alpha}\log(L_n(x_1,\ \ldots,\ x_n;\alpha)) = 0 \Leftrightarrow \alpha = \frac{1}{n}\sum_{i=1}^n (x_i - m_i)\ .$$

L'étude du signe de la dérivée de la log-vraisemblance, montre qu'elle atteint son maximum en $\frac{1}{n}\sum_{i=1}^n (x_i - m_i)$. L'estimateur du maximum de vraisemblance est donc :

$$\hat{\alpha}_n = \frac{1}{n}\sum_{i=1}^n (X_i - m_i)\ .$$

La loi de $\hat{\alpha}_n$ est la loi $\mathcal{N}(\alpha,\ \sigma^2/n)$. En particulier, cet estimateur est sans biais. On vérifie qu'il est efficace. On a également :

$$\frac{\partial^2}{\partial^2\alpha}\log(L_n(x_1,\ \ldots,\ x_n;\alpha)) = -\frac{n}{\sigma^2},$$

et, on en déduit que l'information de Fisher est :

$$I_n = \mathrm{E}_\alpha[-\frac{\partial^2}{\partial^2\alpha}\log(L_n(X_1,\ \ldots,\ X_n;\alpha))] = \frac{n}{\sigma^2}.$$

Comme $\mathrm{Var}_\alpha(\hat{\alpha}_n) = \frac{\sigma^2}{n} = \frac{1}{I_n}$, l'estimateur est efficace.
De plus, les variables $(X_i - m_i)$ sont indépendantes et de même loi gaussienne. Par la loi forte des grands nombres, l'estimateur est convergent. Comme la loi de $\sqrt{n}(\hat{\alpha}_n - \alpha)$ est la loi $\mathcal{N}(0,\ \sigma^2)$, l'estimateur du maximum de vraisemblance est donc asymptotiquement normal (et asymptotiquement efficace).

2. La loi de X_i est la loi $\mathcal{N}(\beta m_i,\ \sigma^2)$. En particulier, la loi de $\hat{\beta}_n$ est la loi $\mathcal{N}(\beta,\ \frac{\sigma^2}{n^2}\sum_{i=1}^n \frac{1}{m_i^2})$. Ainsi $\hat{\beta}_n$ est un estimateur sans biais de β. On vérifie qu'il est efficace. La log-vraisemblance s'écrit :

$$\log(L_n(x_1,\ \ldots,\ x_n;\beta)) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \beta m_i)^2$$

On a :

$$\frac{\partial^2}{\partial^2\beta}\log(L_n(x_1,\ \ldots,\ x_n;\beta)) = -\frac{1}{\sigma^2}\sum_{i=1}^n m_i^2,$$

et donc :

$$I_n = \mathrm{E}_\beta[-\frac{\partial^2}{\partial^2\beta}\log(L_n(X_1,\ \ldots,\ X_n;\beta))] = \frac{1}{\sigma^2}\sum_{i=1}^n m_i^2.$$

D'autre part on a :
$\mathrm{Var}\ (\hat{\beta}_n) = \frac{\sigma^2}{n^2}\sum_{i=1}^n \frac{1}{m_i^2}$.
Par Cauchy-Schwarz, on a $\frac{1}{n^2}\sum_{i=1}^n \frac{1}{m_i^2} \geq \frac{1}{\sum_{i=1}^n m_i^2}$, et l'inégalité est stricte dès qu'il existe $m_i \neq m_j$. En particulier $\mathrm{Var}(\hat{\beta}_n) > \frac{1}{I_n}$, s'il existe $m_i \neq m_j$, et l'estimateur n'est alors pas efficace.

3. On a :

$$\frac{\partial}{\partial\beta}\log(L_n(x_1,\ \ldots,\ x_n;\beta)) = \frac{1}{\sigma^2}\sum_{i=1}^n m_i(x_i - \beta m_i)\ .$$

En étudiant le signe de cette dérivée, on en déduit que l'estimateur du maximum de vraisemblance de β est :

$$\tilde{\beta}_n = \frac{\sum_{i=1}^n m_i X_i}{\sum_{i=1}^n m_i^2}$$

La loi de $\tilde{\beta}_n$ est la loi $\mathcal{N}(\beta,\ \dfrac{\sigma^2}{\sum_{i=1}^n m_i^2})$. En particulier, cet estimateur est sans biais et il est efficace. Il est préférable à $\hat{\beta}_n$.

4. On obtient les estimations avec les intervalles de confiance de niveau exact 95% : $\hat{\alpha}_n \simeq 88.6 \pm 6.1, \hat{\beta}_n \simeq 1.088 \pm 0.006$ et $\tilde{\beta}_n \simeq 1.087 \pm 0.006$. (La théorie des tests permet de déterminer lequel des deux effets (additif ou multiplicatif) modélise au mieux les données observées.)

Exercice 4

1. a) Vérifier que la v.a. $U+V$ est de loi $G(a+b,\lambda)$

   $U \to G(a,\lambda)$, $V \to G(b,\lambda)$

   $\Phi_U(t) = \left(\frac{\lambda}{\lambda-it}\right)^a$

   En effet, $\Phi_U(t) = E\left(e^{itU}\right) = \int_{\mathbb{R}} e^{itu} f_U(u)du = \int_0^{+\infty} e^{itu} \frac{\lambda^a}{\Gamma(a)} u^{a-1} e^{-\lambda u} du = \int_0^{+\infty} \frac{\lambda^a}{\Gamma(a)} u^{a-1} e^{-(\lambda-it)u} du$

   Soit $z = \frac{\lambda-it}{\lambda} u$

   $\Phi_U(t) = \int_0^{+\infty} \frac{\lambda^a}{\Gamma(a)} \left(\frac{\lambda z}{\lambda-it}\right)^{a-1} e^{-\lambda z} \frac{\lambda}{\lambda-it} dz = \left(\frac{\lambda}{\lambda-it}\right)^a \int_0^{+\infty} \frac{\lambda^a}{\Gamma(a)} z^{a-1} e^{-\lambda z} dz = \left(\frac{\lambda}{\lambda-it}\right)^a$

   $\Phi_{U+V}(t) = E\left(e^{it(U+V)}\right) = E\left(e^{itU}e^{itV}\right) = E\left(e^{itU}\right)E(e^{itV})$ (car U et V sont indépendantes)

   $\Phi_{U+V}(t) == \Phi_U(t)\Phi_V(t) = \left(\frac{\lambda}{\lambda-it}\right)^a \left(\frac{\lambda}{\lambda-it}\right)^b = \left(\frac{\lambda}{\lambda-it}\right)^{a+b}$

   $\Longrightarrow U+V \to G(a+b,\lambda)$

   b) Vérifier que que la v.a. cU est de loi $G\left(a, \frac{\lambda}{c}\right)$

   Soit $Z = cU$

   $\Phi_Z(t) = \Phi_{cU}(t) = E\left(e^{itcU}\right) = \Phi_U(ct) = \left(\frac{\lambda}{\lambda-ict}\right)^a = \left(\frac{1}{1-i\left(\frac{c}{\lambda}\right)t}\right)^a$

   $\Longrightarrow Z = cU \to G(a, \frac{c}{\lambda})$

2. $X \to N(0,1)$ et $U \to G(a,\lambda)$.

   Les fonctions de densité des v.a. X et U sont :

   $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\ \forall x \in \mathbb{R}$ et $f_U(u) = \frac{\lambda^a}{\Gamma(a)} u^{a-1} e^{-\lambda u}\ \forall x > 0$

   a) Vérifier que la densité de la v.a. X^2 est de loi $G\left(\frac{1}{2}, \frac{1}{2}\right)$

   Soit $Z = X^2$

   $F_Z(z) = P(Z \leq z) = P(X^2 \leq z) = P(-\sqrt{z} \leq X \leq \sqrt{z}) = 2F_X(\sqrt{z}) - 1$

   $f_Z(z) = \frac{\partial}{\partial z} F_Z(z) = \frac{2}{2\sqrt{z}} f_X(\sqrt{z}) = \frac{1}{\sqrt{z}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z}{2}}\ \forall z > 0$

   $f_Z(z) = \frac{\left(\frac{1}{2}\right)^{\frac{1}{2}}}{\Gamma\left(\frac{1}{2}\right)} z^{\frac{1}{2}-1} e^{-\frac{1}{2}u} \quad \forall z > 0$

   $Z = X^2 \to G\left(\frac{1}{2}, \frac{1}{2}\right)$

b) i. On pose $T = \frac{X}{\sqrt{U}}$ et $S = \sqrt{U}$, Déterminer la densité du couple (T, S)

Soit $f_{T,S}(t,s)$ la densité du couple (T, S)

$\left\{ \begin{array}{l} T = \frac{X}{\sqrt{U}} \\ S = \sqrt{U} \end{array} \right. \Longleftrightarrow \left\{ \begin{array}{l} X = TS \\ U = S^2 \end{array} \right.$

$x \in \mathbb{R} \Longrightarrow t = \frac{x}{\sqrt{u}} \in \mathbb{R}$

$s = \sqrt{u} > 0$

$f_{T,S}(t,s) = f_{X,U}(ts, s^2) \left| \det J \right|$ pour $t \in \mathbb{R}$ et $s > 0$

où $J = \begin{pmatrix} \frac{\partial x(t,s)}{\partial t} & \frac{\partial x(t,s)}{\partial s} \\ \frac{\partial u(t,s)}{\partial t} & \frac{\partial u(t,s)}{\partial s} \end{pmatrix} = \begin{pmatrix} s & t \\ 0 & 2s \end{pmatrix}$

$\left| \det J \right| = 2s^2$

X et U deux variables aléatoires réelle indépendantes donc $f_{X,U}(ts, s^2) = f_X(ts) f_U(s^2)$

$f_{T,S}(t,s) = f_{X,U}(ts, s^2) \left| \det J \right| = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2 s^2}{2}} \frac{\lambda^a}{\Gamma(a)} \left(s^2\right)^{a-1} e^{-\lambda s^2} 2s^2 = \frac{2\lambda^a}{\sqrt{2\pi}\Gamma(a)} s^{2a} e^{-(\lambda + \frac{t^2}{2})s^2}$

pour $t \in \mathbb{R}$ et $s > 0$

b) ii. Déterminer la densité de la v.a. T.

$f_T(t)$ est la densité marginale de la v.a. T

$f_T(t) = \int_{\mathbb{R}} f_{T,S}(t,s) ds = \int_0^{+\infty} \frac{2\lambda^a}{\sqrt{2\pi}\Gamma(a)} s^{2a} e^{-(\lambda + \frac{t^2}{2})s^2} ds \overset{z=s^2}{=} \int_0^{+\infty} \frac{2\lambda^a}{\sqrt{2\pi}\Gamma(a)} z^a e^{-(\lambda + \frac{t^2}{2})z} \frac{1}{2\sqrt{z}} dz = \int_0^{+\infty} \frac{\lambda^a}{\sqrt{2\pi}\Gamma(a)} z^{a-\frac{1}{2}} e^{-(\lambda + \frac{t^2}{2})z} dz$

$= \frac{\lambda^a}{\sqrt{2\pi}\Gamma(a)} \int_0^{+\infty} z^{\left(a+\frac{1}{2}\right)-1} e^{-(\lambda + \frac{t^2}{2})z} dz = \frac{\lambda^a}{\sqrt{2\pi}\Gamma(a)} \frac{\Gamma(a+\frac{1}{2})}{\left(\lambda + \frac{t^2}{2}\right)^{a+\frac{1}{2}}} = \frac{\lambda^a}{\sqrt{2\pi}\Gamma(a)} \frac{\Gamma(a+\frac{1}{2})}{\lambda^{a+\frac{1}{2}} \left(1 + \frac{t^2}{2\lambda}\right)^{a+\frac{1}{2}}} = \frac{\Gamma(a+\frac{1}{2})}{\sqrt{\lambda}\sqrt{2\pi}\Gamma(a)} \frac{1}{\left(1 + \frac{t^2}{2\lambda}\right)^{a+\frac{1}{2}}}$

$\Longrightarrow f_T(t) = \frac{\Gamma(a+\frac{1}{2})}{\sqrt{\lambda}\sqrt{2\pi}\Gamma(a)} \frac{1}{\left(1 + \frac{t^2}{2\lambda}\right)^{a+\frac{1}{2}}} \quad \forall t \in \mathbb{R}.$

3. déduire sans calcul à partir de 1.b) et 2.b) la densité de la v.a. $X_1^2 + ... + X_n^2$

   $X \to N(0,1)$ alors $X^2 \to G\left(\frac{1}{2}, \frac{1}{2}\right)$

   $U \to G(a, \lambda)$, $V \to G(b, \lambda)$ alors $U + V \to G(a + b, \lambda)$ et $cU \to G(a, \frac{c}{\lambda})$

   Donc $X_i^2 \to G(\frac{1}{2}, \frac{1}{2})$ et $X_1^2 + ... + X_n^2 \to G(\frac{n}{2}, \frac{1}{2})$

4. déduire sans calcul la densité de la v.a. $\frac{Y}{\sqrt{\frac{X_1^2 + ... + X_n^2}{n}}}$ (loi de Student)

   $U \to G(a, \lambda)$ alors $cU \to G(a, \frac{c}{\lambda})$

   $X_1^2 + ... + X_n^2 \to G(\frac{n}{2}, \frac{1}{2}) \Longrightarrow \frac{1}{n}\left(X_1^2 + ... + X_n^2\right) \to G(\frac{n}{2}, \frac{n}{2})$

   La densité de $T = \frac{X}{\sqrt{U}}$ où $X \to N(0,1)$ et $U \to G(a, \lambda)$ est $f_T(t) = \frac{\Gamma(a+\frac{1}{2})}{\sqrt{\lambda}\sqrt{2\pi}\Gamma(a)} \frac{1}{\left(1 + \frac{t^2}{2\lambda}\right)^{a+\frac{1}{2}}} \quad \forall t \in \mathbb{R}$

   La densité de la variable $\frac{Y}{\sqrt{\frac{X_1^2 + ... + X_n^2}{n}}}$ où $Y \to N(0,1)$ et $\frac{X_1^2 + ... + X_n^2}{n} \to G(\frac{n}{2}, \frac{n}{2})$

   est la densité de $T = \frac{X}{\sqrt{U}}$ avec $a = \frac{n}{2}$ et $\lambda = \frac{n}{2}$

$$f_{\frac{Y}{\sqrt{\frac{X_1^2+\ldots+X_n^2}{n}}}}(t) = \frac{\Gamma(\frac{n}{2}+\frac{1}{2})}{\sqrt{\frac{n}{2}}\sqrt{2\pi}\Gamma(\frac{n}{2})}\frac{1}{\left(1+\frac{t^2}{n}\right)^{\frac{n}{2}+\frac{1}{2}}} = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n}\Gamma(\frac{n}{2})\Gamma(\frac{1}{2})}\frac{1}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}} \quad \forall t \in \mathbb{R}$$

Exercice 5

1. On associe à chaque usine une v.a.r $X_i = 1$ si l'usine i respecte les NHS et 0 sinon.

   $X_i \to B(p)$ où $p = 0.4$.
   Soit f_n la fréquence des usines qui respectent les NHS dans un échantillon de n usines.
   $f_n = \frac{\sum_{i=1}^n X_i}{n}$
   On cherche n tel que $P(0.35 \leq f_n \leq 0.45) \geq 0.9$
   D'après le TCL, $\frac{f_n - E(f_n)}{\sqrt{V(f_n)}} \underset{n\to\infty}{\overset{L}{\to}} N(0,1)$
   $E(f_n) = E\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{1}{n}E\left(\sum_{i=1}^n X_i\right) = \frac{1}{n}\sum_{i=1}^n E(X_i) = \frac{1}{n}\times n\times E(X_i) = p$
   $V(f_n) = V\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{1}{n^2}V\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\sum_{i=1}^n V(X_i) = \frac{1}{n^2}\times n\times V(X_i) = \frac{p(1-p)}{n}$
   $f_n^* = \frac{f_n - E(f_n)}{\sqrt{V(f_n)}} = \frac{f_n - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{f_n - 0.4}{\sqrt{\frac{0.24}{n}}} \underset{n\to\infty}{\overset{L}{\to}} N(0,1)$
   $P(0.35 \leq f_n \leq 0.45) \geq 0.9 \iff P\left(\frac{0.35-0.4}{\sqrt{\frac{0.24}{n}}} \leq \frac{f_n-0.4}{\sqrt{\frac{0.24}{n}}} \leq \frac{0.45-0.4}{\sqrt{\frac{0.24}{n}}}\right) \geq 0.9$
   $P\left(\frac{-0.05}{\sqrt{\frac{0.24}{n}}} \leq f_n^* \leq \frac{0.05}{\sqrt{\frac{0.24}{n}}}\right) \geq 0.9 \iff P\left(|f_n^*| \leq \frac{0.05}{\sqrt{\frac{0.24}{n}}}\right) \geq 0.9$
   $2F_{N(0,1)}\left(\frac{0.05}{\sqrt{\frac{0.24}{n}}}\right) - 1 \geq 0.9 \iff F_{N(0,1)}\left(\frac{0.05}{\sqrt{\frac{0.24}{n}}}\right) \geq 0.95$

   $\frac{0.05}{\sqrt{\frac{0.24}{n}}} \geq F_{N(0,1)}^{-1}(0.95) \iff \sqrt{\frac{0.24}{n}} \leq \frac{0.05}{1.64} \iff n \geq 258.2$
   Il faut contrôler au moins un échantillon de 259 usines pourque la fréquence des usines qui respectent les NHS (parmi toutes les usines) soit comprise entre 35% et 45% avec une probabilité supérieure à 0.9. Ou en d'autres termes, il faut prendre un échantillon d'au moins 259 usines pour pouvoir affirmer avec un risque d'erreur au plus égal à 10% que la fréquence des usines qui respectent les NHS se situe entre 35% et 45%.

2. On contrôle 128 usines et on observe que 72 d'entre elles respectent les normes d'hygiène et de sécurité. Trouver un intervalle de confiance de niveau 95% pour la proportion d'usines respectant les normes d'hygiène.
   $f_n = \frac{\sum_{i=1}^n X_i}{n} = \frac{72}{128} = 0.5625$
   Le niveau de confiance est $1 - \alpha = 0.95$
   Le seuil de risque est $\alpha = 0.05$
   D'après le TCL,
   $f_n^* = \frac{f_n - p}{\sqrt{\frac{p(1-p)}{n}}} \underset{n\to\infty}{\overset{L}{\to}} N(0,1)$

$P\left(\left|\frac{f_n-p}{\sqrt{\frac{p(1-p)}{n}}}\right| \leq t_{1-\frac{\alpha}{2}}\right) = 1-\alpha \Longleftrightarrow P\left(f_n - t_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}} \leq p \leq f_n + t_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}\right) = 1-\alpha$

$P\left(p \in \left[f_n - t_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}\ ,\ f_n + t_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}\right]\right) = 1-\alpha$

$IC_{1-\alpha}(p) = \left[f_n - t_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}\ ,\ f_n + t_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}\right]$

On décide d'estimer $p(1-p)$ par $f_n(1-f_n)$ dans $IC_{1-\alpha}(p)$ (voir la méthode exacte et la méthode par majoration de la variance dans le cours), on a alors :

$f_n = 0.5625$

$n = 128$

$\hat{\sigma}_{f_n} = \sqrt{\frac{f_n(1-f_n)}{n}} = \sqrt{\frac{0.5625(1-0.5625)}{128}} = 4.3848 \times 10^{-2}$

$t_{1-\frac{\alpha}{2}} = t_{0.975}$

$P\left(|f_n^*| \leq t_{1-\frac{\alpha}{2}}\right) = 1-\alpha \Longleftrightarrow 2F_{N(0,1)}\left(t_{1-\frac{\alpha}{2}}\right) - 1 = 1-\alpha \Longleftrightarrow F_{N(0,1)}\left(t_{1-\frac{\alpha}{2}}\right) = 1-\frac{\alpha}{2} \Longleftrightarrow t_{1-\frac{\alpha}{2}} = F^{-1}_{N(0,1)}\left(1-\frac{\alpha}{2}\right)$

$t_{0.975} = F^{-1}_{N(0,1)}(0.975) = 1.96$

$IC_{95\%}(p) = \left[0.5625 - 1.96 \times \sqrt{\frac{0.5625(1-0.5625)}{128}}\ ,\ 0.5625 + 1.96 \times \sqrt{\frac{0.5625(1-0.5625)}{128}}\right] = [0.4766\ , 0.6484]$

D'après l'échantillon de 128 usines qu'on a contrôlées, on est sûr à 95% que la proportion des usines qui respectent les NHS se situe entre 47.66% et 64.84%.

Exercice 6

1. La statistique qu'on utilise pour trouver un intervalle de confiance pour la variance σ^2 est la statistique du Khi-Deux. Il y a deux cas : cas où la moyenne m est connue et cas où la moyenne m est inconnue. Ici la moyenne m est inconnue.

   Soit S^2 la variance empirique de l'échantillon (l'estimateur sans biais de σ^2) : $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$

   $\frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma^2}\right)^2 \to \chi^2(n-1)$

   $P\left(\chi^2_{\frac{\alpha}{2},n-1} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{1-\frac{\alpha}{2},,n-1}\right) = 1-\alpha$

   où

   $\chi^2_{\frac{\alpha}{2}}$ est telle que $P\left(\chi^2 \leq \chi^2_{\frac{\alpha}{2},n-1}\right) = \frac{\alpha}{2}$

   $\chi^2_{1-\frac{\alpha}{2}}$ est telle que $P\left(\chi^2 \leq \chi^2_{1-\frac{\alpha}{2},n-1}\right) = 1-\frac{\alpha}{2}$

   On aura $P\left(\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},,n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}}\right) = 1-\alpha$

   $P\left(\sigma^2 \in \left[\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},,n-1}}, \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}}\right]\right) = 1-\alpha$

$IC_{1-\alpha}(\sigma^2) = \left[\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},,n-1}}, \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}}\right]$

AN :

Le niveau de confiance est $1-\alpha = 95\% \Longleftrightarrow$ le niveau du risque est 5%

$n = 26$

$P\left(\chi^2 \leq \chi^2_{\frac{\alpha}{2},n-1}\right) = \frac{\alpha}{2} \Longrightarrow P\left(\chi^2 \leq \chi^2_{2.5\%,25}\right) = 2.5\% \Longrightarrow \chi^2_{2.5\%,25} = 13.12$

$P\left(\chi^2 \leq \chi^2_{1-\frac{\alpha}{2},n-1}\right) = 1-\frac{\alpha}{2} \Longrightarrow P\left(\chi^2 \leq \chi^2_{97.5\%,25}\right) = 97.5\% \Longrightarrow \chi^2_{97.5\%,25} = 40.642$

$IC_{95\%}(\sigma^2) = \left[\frac{1.132}{40.642}, \frac{1.132}{13.12}\right] = [0.0279\ ,\ 0.0863]$

2. Pour trouver un intervalle de confiance pour m on utilise la statistique de la loi normale centrée et réduite si σ est connu et la statistique de Student à $n-1$ degrés de liberté si σ est inconnu. Dans ce cas σ est inconnu.

   $T = \frac{\bar{X}_n - m}{\frac{S}{\sqrt{n}}} \to t(n-1)$

   $P(|T| \leq t) = 1-\alpha$

   $P\left(\left|\frac{\bar{X}_n - m}{\frac{S}{\sqrt{n}}}\right| \leq t\right) = 1-\alpha$

   $P\left(\bar{X}_n - t\frac{S}{\sqrt{n}} \leq m \leq \bar{X}_n + t\frac{S}{\sqrt{n}}\right) = 1-\alpha$

   $P\left(m \in \left[\bar{X}_n - t\frac{S}{\sqrt{n}}\ , \bar{X}_n + t\frac{S}{\sqrt{n}}\right]\right) = 1-\alpha$

   $IC_{1-\alpha}(m) = \left[\bar{X}_n - t\frac{S}{\sqrt{n}}\ , \bar{X}_n + t\frac{S}{\sqrt{n}}\right]$

   AN :

   $1-\alpha = 90\% \Longleftrightarrow \alpha = 10\%$

   $\bar{X}_n = 62$

   $n = 26$

   $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2} = \sqrt{\frac{1}{26-1} \times 1.132} = 0.2128$

   $P(|T| \leq t) = 0.9 \Longleftrightarrow 2P(T \leq t) - 1 = 0.9 \Longrightarrow t = 1.708$ (d'après la table de la fonction de répartition d'une Student à 25 degrés de libertés)

   $IC_{90\%}(m) = \left[62 - 1.708 \times \frac{0.2128}{\sqrt{26}}\ , 62 + 1.708 \times \frac{0.2128}{\sqrt{26}}\right] = [61.928\ ,\ 62.071]$

   D'après l'échantillon des 26 pièces prélevées, on est sûr à 90% (niveau de confiance) que la diamètre moyen des pièces produits se situe dans l'intervalle [61. 928 , 62. 071].